



Additional Metadata Guidelines to Improve the Structure and Usability of HDF and NetCDF Files

Sean Leavor – SSAI/NASA LaRC

Megan Buzanowicz – SSAI/NASA LaRC, Matt Tisdale –
NASA LaRC,

Michael Shook – NASA LaRC, Gao Chen – NASA LaRC

HDF and NetCDF File Format

- Useful for handling large volumes of data
- Able to host extensive metadata
- Frequently used with remote sensing data
 - Has been used to support numerous satellite, airborne, and ground-based measurements
- Climate and Forecast (CF) conventions give a structure for data to follow
 - Aim for self-describing data that can be read by both humans and software

Airborne HDF and NetCDF Files

- Amount and content of metadata can vary wildly between campaigns or principal investigators
 - Current guidelines are not always specific enough about what should be included with the metadata
- The Measurements of Aerosols, Clouds, and their Interactions for ESMs (MACIE) group has started an effort to build upon existing metadata conventions
 - Would help to ensure airborne data is better able to support their use for research

FAIR Principles

- Findable
 - Metadata and data should be easy to find for both humans and computers
- Accessible
 - Once found, the user needs to know how data can be accessed
- Interoperable
 - Vocabulary describing the data allows it to be integrated with other data or applications for analysis, storage, or processing
- Reusable
 - Metadata and data should be well-described so that it can be easily understood or replicated for future reuse

Goals

Primary goal is to enhance the interoperability and reusability of data. This can be achieved by:

- Building upon current CF conventions
- Improving the metadata structure and the way it is presented
- Controlling domain relevant information
- Making sure variables are dimension scaled
- Ensuring data use standardized units
- Making arrangements for standardized uncertainty reporting

Common Airborne HDF/NetCDF Metadata Issues

- Data does not have a predictable structure
- Descriptive information about the file can be missing
- Global attributes or variable information may not be present
- Variable dimensions might not be labeled

File "owlets-Car1_Mobile-insitu_20170708_RB.H5"

File type: Hierarchical Data Format, version 5

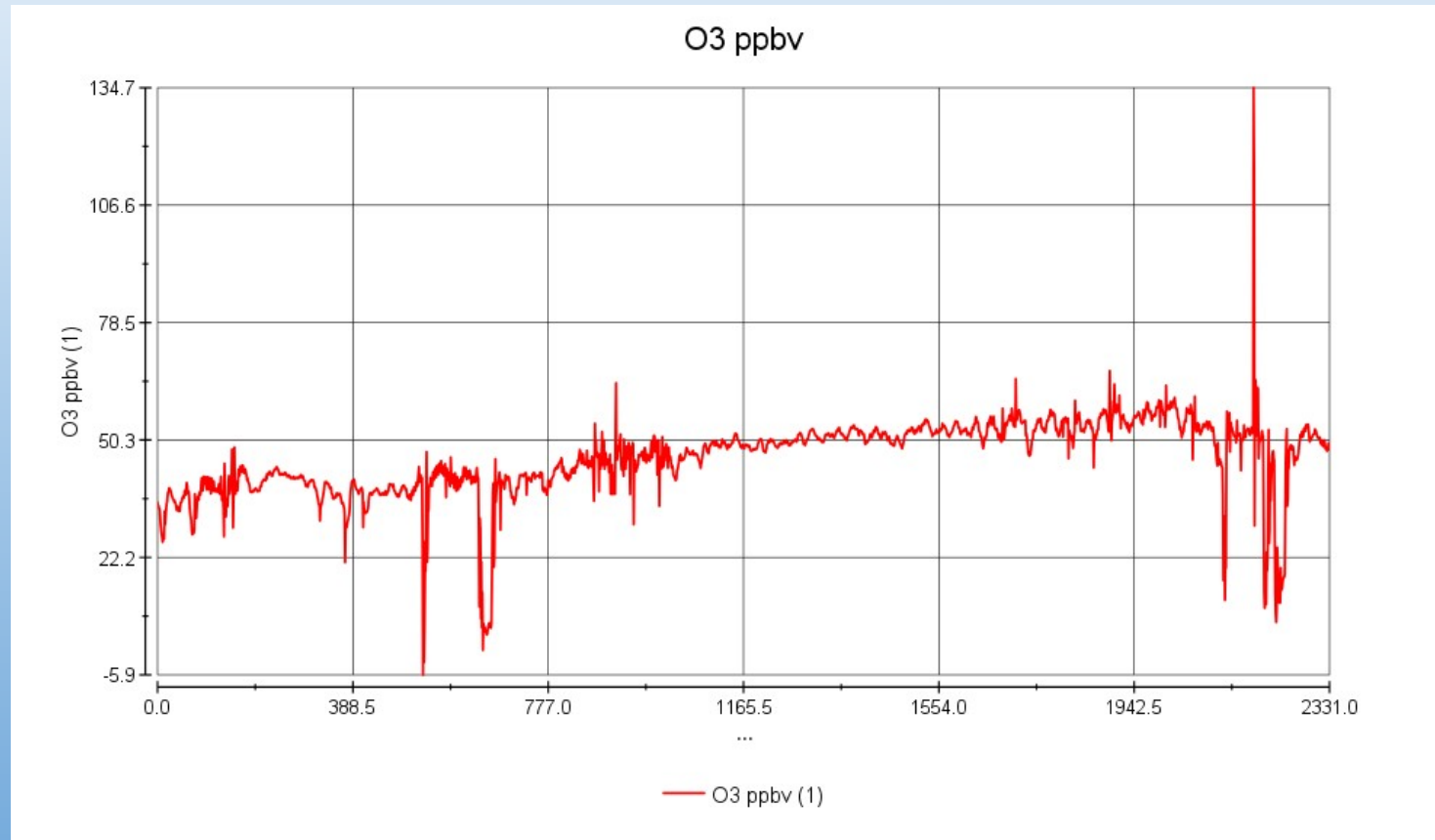
```
netcdf file:/C:/Users/sleavor/Downloads/owlets-Car1_Mobile-insitu_20170708_RB.H5 {  
  variables:  
    String 001_Readme;  
  
    char DateTime.UTC(2332, 19);  
  
    double Latitude_Deg(2332);  
  
    double Longitude_Deg(2332);  
  
    double O3_ppbv(2332);  
  
    double UnixSec.UTC(2332);  
  
}
```

Variable "Latitude_Deg"

In file "owlets-Car1_Mobile-insitu_20170708_RB.H5"

```
double Latitude_Deg(2332);
```

Common Airborne HDF/NetCDF Metadata Issues



Proposed Guidelines

- Standardize file structure based upon CF conventions
- Use global attributes to host required general metadata information
- Require use of CF global and variable attributes for domain relevant metadata
- Indicate linked variables using CF attributes
- Recommend using standardized variables
- Ensure data variables are properly attached to dimensions
- Allow for guidelines to be expanded upon to cater to specific types of data

Proposed Guidelines

Variable "1064_bsc_cloud_screened"

In file "traceraq-HSRL2_GV_20210901_R0.h5"

Variable full name: DataProducts/1064_bsc_cloud_screened

```
double 1064_bsc_cloud_screened(1603, 668);  
:Description = "1064 nm aerosol backscatter coefficient cloud screened";  
:Description2 = "None";  
:Horizontal_resolution_(seconds) = 10.0; // double  
:Vertical_resolution_(meters) = 15.0; // double  
:chi = 1.0; // double  
:molecular_depolarization = 0.0037; // double  
:units = "km^-1sr^-1";  
:code_date = "31-Mar-2022 16:38:54";  
:run_date = "06-Apr-2022 15:32:18";  
:StandardName = "AerOpt_BackScattering_Profile_IR_RHa_Bulk_AMB";
```

Proposed Guidelines

```
// global attributes:
:PI_NAME = "Newchurch;Michael J.";
:PI_AFFILIATION = "University of Alabama Huntsville;UAH";
:PI_ADDRESS = "University of Alabama Huntsville, Atmospheric Science Dept.;Huntsville, AL 35806;UNITED STATES";
:PI_EMAIL = "mike@nsstc.uah.edu";
:DO_NAME = "Kuang;Shi";
:DO_AFFILIATION = "University of Alabama Huntsville;UAH";
:DO_ADDRESS = "University of Alabama Huntsville, Atmospheric Science Dept.;Huntsville, AL 35806;UNITED STATES";
:DO_EMAIL = "kuang@nsstc.uah.edu";
:DS_NAME = "Newchurch;Michael J.";
:DS_AFFILIATION = "University of Alabama Huntsville;UAH";
:DS_ADDRESS = "University of Alabama Huntsville, Atmospheric Science Dept.;Huntsville, AL 35806;UNITED STATES";
:DS_EMAIL = "mike@nsstc.uah.edu";
:DATA_DESCRIPTION = "Tropospheric ozone profile from lidar at Huntsville, AL, US";
:DATA_DISCIPLINE = "ATMOSPHERIC.CHEMISTRY;REMOTE.SENSING;GROUNDBASED";
:DATA_GROUP = "EXPERIMENTAL;PROFILE.STATIONARY";
:DATA_LOCATION = "HUNTSVILLE.AL";
:DATA_SOURCE = "LIDAR.O3_UAH001_HIRES";
:DATA_VARIABLES = "LATITUDE.INSTRUMENT;LONGITUDE.INSTRUMENT;ALTITUDE.INSTRUMENT;DATETIME;DATETIME.START;DATETIME.STOP;INTEGRATION.TIME;ALTITUDE";
:DATA_START_DATE = "20220429T143901Z";
:DATA_STOP_DATE = "20220429T154147Z";
:DATA_FILE_VERSION = "001";
:DATA_MODIFICATIONS = "FINAL";
:DATA_CAVEATS = "refer to http://nsstc.uah.edu/atmchem/lidar/DIAL\_data.html";
```

Summary

- Metadata for airborne HDF and NetCDF files can vary greatly between campaigns or PIs
- Additional guidelines for the metadata structure will make data more interoperable and reusable
 - Can aid data users in their use of measurements for research or for allowing data to be more easily ingested for archival
- Updated standards may need to be catered to individual types of measurements
- Proposed guidelines have been tested for several instruments
 - High Spectral Resolution Lidar (HSRL)
 - High Altitude Laser Observatory (HALO)
 - Research Scanning Polarimeter (RSP)
 - Geostationary Coastal and Air Pollution Event (GEO-CAPE) Airborne Simulator (GCAS)